

CLAIMS

What is claimed is:

1. A method of aggregating data comprising the steps of:
  - 5 a) receiving a plurality of web log records; and
  - b) generating a volume cube having a plurality of dimensions based on the plurality of web log records; and
  - c) generating a first summary cube at a first level of abstraction based on the volume cube; wherein the first summary cube has a plurality of  
10 cells and at least one parent dimension for each dimension of the volume cube; wherein each cell includes a value that represents a summarization of corresponding cell values of the volume cube aggregated over all the dimensions of the volume cube.
- 15 2. The method of claim 1 further comprising:
  - d) generating a second summary cube at a second level of abstraction based on the first summary cube; wherein the second summary cube has at least one parent dimension for each dimension of the first summary cube;
  - 20 wherein the second summary cube is generated by limiting the dimensions of the first summary cube to a range of values above a predetermined value.
3. The method of claim 1 further comprising:
  - 25 d) selectively generating a partial aggregate along one or more dimensions by using a query that specifies at least one dimension along which the partial aggregate is to be calculated.

4. The method of claim 3 wherein the first summary cube includes aggregated values; and wherein the first summary cube includes a drill-down enabling mechanism for specifying relationships between dimensions of the volume cube and the dimensions of the first summary cube for enabling drill-down analysis.
5. The method of claim 1 wherein the first summary cube includes a plurality of dimensions; and wherein each dimension corresponds to at least one dimensions of the volume cube.
6. The method of claim 5 wherein the first summary cube includes a first dimension for specifying an origin; a second dimension for specifying a subject; and a third dimension for specifying a referring site.
7. The method of claim 2 wherein the second summary cube includes a plurality of dimensions; and wherein each dimension corresponds to at least one dimensions of the first summary cube.
8. The method of claim 7 wherein the second summary cube includes a first dimension for specifying a top origin; a second dimension for specifying a top subject; and a third dimension for specifying a top referring site.
9. The method of claim 1 wherein each web log record includes a plurality of fields.
10. The method of claim 9 wherein each web log record includes a first field for storing an IP address of an origin site;

a second field for storing a target URI;  
a third field for storing a referring site; and  
a fourth field for storing a time.

- 5 11. A method for analyzing web access comprising the steps of:
- a) receiving a plurality of web log records; and
  - b) generating multi-dimensional summary information based on the  
plurality of web log records; and
  - c) performing derivation and analysis to discover usage patterns and rules  
10 for supporting business intelligence by using the multi-dimensional  
summary information.
12. The method of claim 11 wherein the step of performing derivation and  
analysis to discover usage patterns and rules for supporting business  
15 intelligence includes one of usage analysis, web site traffic analysis, and  
business rules discovery.
13. The method of claim 11 further comprising the steps of:
- d) performing multilevel and multidimensional analysis on the summary  
20 information to generate one of multidimensional patterns,  
multilevel patterns, and probability based patterns.
14. The method of claim 11 further comprising the steps of:
- d) generating multilevel and multidimensional feature ranking cubes for  
25 ranking web access along multiple dimensions and at multiple  
levels.

15. The method of claim 14 wherein the step of generating multilevel and multidimensional feature ranking cubes for ranking web access along multiple dimensions and at multiple levels includes:

5 generating a first cube for ranked list of elements of a particular dimension, where a feature is represented by a dimension; and generating a second cube for one of volume and probability distribution corresponding to the ranked list of elements of a particular dimension.

- 10 16. The method of claim 11 further comprising the steps of:  
d) performing correlation analysis on the summary information to generate association rules for use in web access analysis.

- 15 17. The method of claim 16 wherein the step of performing correlation analysis on the summary information to generate association rules for use in web access analysis includes generating one of multilevel association rules with flexible base and dimensions and time-variant association rules.

18. A system for analyzing web access comprising:  
20 a) a source of web log records; and  
b) an OLAP engine when executing a web access analysis program for receiving a plurality of web log records, generating multi-dimensional summary information based on the web log records; and performing derivation and analysis to discover usage patterns  
25 and rules for supporting business intelligence by using the multi-dimensional summary information.

19. The system of claim 18 wherein the web access analysis program includes

a feature ranking facility for generating multilevel and multidimensional feature ranking cubes for ranking web access along multiple dimensions and at multiple levels; wherein the feature ranking facility generates a first cube for ranked list of elements of a particular dimension, where a feature is represented  
5 by a dimension; and generates a second cube for one of volume and probability distribution corresponding to the ranked list of elements of a particular dimension.

20. The system of claim 18 wherein the web access analysis program includes  
10 a correlation analysis facility for performing correlation analysis on the summary information to generate association rules for use in web access analysis; wherein the correlation analysis facility generates one of multilevel association rules with flexible base and dimensions and time-variant association rules.

15 21. The system of claim 18 wherein the web access analysis program includes a direct binning facility for concurrently generating a volume cube based on the plurality of web log records and directly generating a high diagonal cube based on plurality of web log records.